Hessisches Kompetenzzentrum
für Hochleistungsrechnen

Deep Reinforcement Learning for Robotics



Project Manager
Dr. Davide Tateo

Researchers
Puze Liu, Firas Al-Hafez and Bane Janjus

Principal Investigator
Dr. Davide Tateo

Project Term
2022 - 2023

Clusters
Lichtenberg II Cluster Darmstadt

Software
Python, PyTorch

Institute
Intelligent Autonomous Systems

University
Technische Universität Darmstadt

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN
TECHNISCHE UNIVERSITÄT DARMSTADT
UNIKASSEL VERSITÄT
Philipps Universität Marburg

Introduction

Robotics platforms can massively benefit from novel Deep Reinforcement Learning approaches. However, robotics have different requirements from a standard AI reinforcement learning environment, as the algorithm should run in the real world. Furthermore, we can exploit domain knowledge that cannot be assumed for a general AI agent, such as robot kinematics. In this project, we want to investigate how we can adapt deep reinforcement learning approaches to the field of robotics. In particular, we will focus on three main challenges:

- The first challenge is safety, as the robot should not be allowed to cause damage to itself, the environment, or other people.
- The second is the definition of the reward function, as it is often difficult to properly define the performance metric for a real robotic platform. Indeed, the behavior can be very difficult to specify (for example in the locomotion scenario) or the reward function that specifies the desired behavior may be too sparse to be effectively optimized.
- The third challenge is interpretability, as black-box neural networks cannot be trusted in general to be deployed in the real world.

To solve robotics tasks, we use Deep Reinforcement learning approaches, e.g. DQN, PPO, TRPO, and SAC algorithms, and Imitation Learning methods, such as Inverse Reinforcement Learning techniques, e.g. VAIL, GAIL, and IQ-Learn. All these approaches are extremely data-hungry and we require many runs in simulation to identify the algorithm hyperparameters that maximize the learning performance. Therefore, it is fundamental to run extensive experimentation on a high-performance computing system. The parallel execution allows us to test the algorithm multiple times, reducing the variability of experiments due to the stochasticity of the policies and the learning algorithms. In particular, these approaches are fundamental in the setting of interpretable reinforcement learning, where genetic algorithms are used to extract interpretable policies. These algorithms rely on parallel execution and random search, making it particularly relevant to work with a high-performance computing system.

Methods

We have developed three different methodologies facing the three different challenges highlighted by the project:

- To impose safety constraints on deep reinforcement learning algorithms, we developed a technique that allows us to generate a safe action space, where every action sampled is safe. This is achieved by building the so-called “constraint manifold” and exploiting the tangent space of this constraint manifold to build the safe action space.
- To improve imitation learning performance, we analyzed the mathematical formulation of inverse reinforcement learning regularizers. This led to a different formulation of the Inverse Reinforcement Learning problem allowing us to frame it as a fixed reward setting. In this setting, it is easy to use heuristics to

JUSTUS-LIEBIG-UNIVERSITÄT GIESSEN TECHNISCHE UNIVERSITÄT DARMSTADT UNIKASSEL VERSITÄT Philipps Universität Marburg

boost the learning process.
- To generate the interpretable policies, we mix the DQN algorithm with generic programming, allowing us to learn tree-based policies.

Results

We achieved very good results in the context of safety. Our methods have been proven to generalize in many different tasks, such as navigation manipulation, and interaction with humans. The methodology developed allows the robot to act and explore the environment while respecting safety constraints. In the long term, this approach will allow to deploy learning robots in the real world, preventing damage to the environment, people, or the robot itself.
In terms of learning without a reward function, we performed an empirical and practical analysis of the current state of the art. Thanks to our experiments we identified some key issues of current state-of-the-art Inverse Reinforcement Learning methods, and we developed an algorithm that massively boosts the stability of the learning process, allowing us to solve extremely complex locomotion tasks involving complex humanoid kinematics.
In terms of Interpretable Reinforcement learning, we have only minor results. Unfortunately, learning interpretable structures for complex tasks appears to be still complicated, and naive methods are not able to tackle the complexity. The main issue lies in the instability of Reinforcement Learning methods under off-policy distributions. This problem is well-known in the literature, but already-known stabilization techniques are not able to deal with interpretable structures such as trees.

Discussion

While we have made good progress towards the objective of learning in the robotics domain, we are still far from a comprehensive theory and generalized algorithmic structure that allows us to tackle the problem completely and deploy learning agents in the real world. We believe, however, that the progress in the field of Safe Reinforcement Learning and Inverse Reinforcement learning will be one of the stepping stones to build the robotics applications of the future.

Figures



Figure 1: Real Tiago robot safely interacts in a real-world environment, while avoiding collision with complex geometries

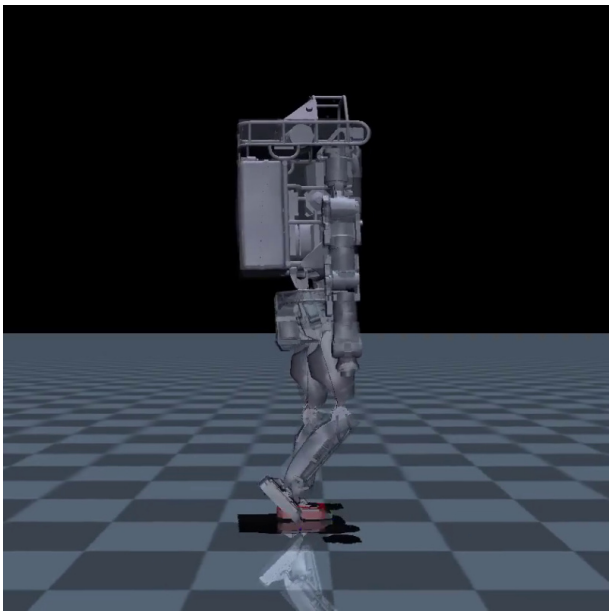

Figure 2: Simulated Atlas robot learning to walk from demonstrated data.

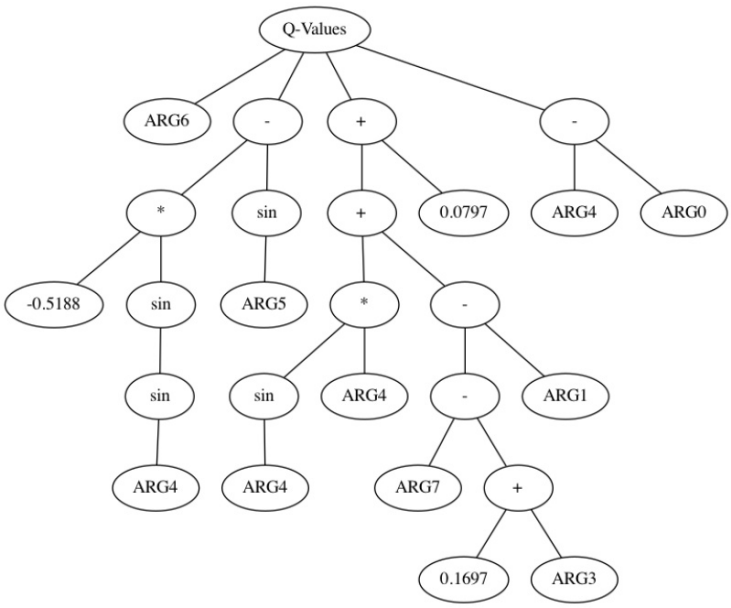

Figure 3: Example of interpretable policy learned during the interaction with the environment.

Publications

Liu, P; Zhang, K; Tateo, D; Jauhri, S; Hu, Z; Peters J; Chalvatzaki, G. Safe Reinforcement Learning of Dynamic High-Dimensional Robotic Tasks: Navigation, Manipulation, Interaction. ICRA, London, UK (2023) https://doi.org/10.1109/ICRA48891.2023.10161548

Al-Hafez, F; Tateo, D; Arenz, O; Zhao, G; Peters, J. LS-IQ: Implicit Reward Regularization For Inverse Reinforcement Learning. ICLR, Kigali, Rwanda (2023) https://doi.org/10.48550/arXiv.2303.00599

Last Update: 2023-08-02 14:13